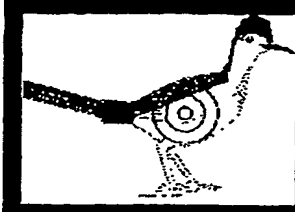


FIG. 1


1/40

File Edit View Go Communicator

Bookmarks Location: <http://victoria.inpharmatica.co.uk/~volker/BPD3target.html> WIPO



Target Mining Interface



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):
- OR
- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

Filter for the following SPECIES:

<input type="checkbox"/> Homo sapiens	<input type="checkbox"/> Rattus norvegicus (Rat)	<input type="checkbox"/> Mus musculus (Mouse)	<input type="checkbox"/> Danio rerio (Zebra fish)
---------------------------------------	--	---	---

100%

FIG. 2A

2/40

File Edit View Go Communicator Help													
Bookmarks Location http://London-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl													
2) 84 additional hits identified by both, Genome Threader and PSI-BLAST:													
Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!													
Add2list	EPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st iter. (PSI)	Best iter. (PSI)	Ex- v (P)
J	AAAS9544.1 drill through Top50BlastHits Red Seq. View	AAAS9544.1	Not given	Homo sapiens	PRI	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E
J	AAB24821.1 drill through Top50BlastHits Red Seq. View	AAB24821.1	leukocyte integrin alpha chain	Homo sapiens	PRI	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E
J	Q99715 drill through Top50BlastHits Red Seq. View	Q99715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	28.6%, 26% unmaskedSW	4-181, 2-174	140-318, 2321-2495	440	100% unmaskedGT	1	3	2E
J	P20701 drill through Top50BlastHits Red Seq. View	P20701	LEUKOCYTE ADHESION GLYCOPROTEIN LFA-1 ALPHA CHAIN PRECURSOR (LEUKOCYTE FUNCTION ASSOCIATED MOLECULE 1, ALPHA CHAIN) (CD11A) (INTEGRIN ALPHA-L).	Homo sapiens (Human).	PRI	100%, 100% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E
J	AAC31672.1 drill through Top50BlastHits Red Seq. View	AAC31672.1	leukocyte function-associated molecule-1 alpha subunit	Homo sapiens	PRI	99.5%, 99% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E
J	CAA72402.1 drill through Top50BlastHits Red Seq. View	CAA72402.1	collagen type XIV	Homo sapiens	PRI	29.1%, 29% unmaskedSW	2-180, 2-180	5-185, 5-185	422	100% unmaskedGT	1	2	2E
J	AAB38702.1 drill through Top50BlastHits Red Seq. View	AAB38702.1	cartilage matrix protein	Homo sapiens	PRI	31.7%, 27% unmaskedSW	4-183, 2-182	275-455, 39-223	413	100% unmaskedGT	1	2	2E
J	CAB70853.1 drill through Top50BlastHits Red Seq. View	CAB70853.1	hypothetical protein	Homo sapiens	PRI	28%, 28% unmaskedSW	1-180, 1-183	437-620, 437-624	406	100% unmaskedGT	1	2	1E
J	CAA27972.1 drill through Top50BlastHits Red Seq. View	CAA27972.1	Not given	Homo sapiens	PRI	20.5%, 20% unmaskedSW	3-183, 2-181	1497-1673, 1689-1873	405	100% unmaskedGT	2	3	1E
J	AAB59512.1 drill through Top50BlastHits Red Seq. View	AAB59512.1	Not given	Homo sapiens	PRI	20.5%, 20% unmaskedSW	3-183, 2-181	758-934, 950-1134	405	100% unmaskedGT	2	3	1E
J	CAA07569.1 drill through Top50BlastHits Red Seq. View	CAA07569.1	matrin-4	Homo sapiens	PRI	28.1%, 25% unmaskedSW	1-183, 1-183	342-528, 31-217	403	100% unmaskedGT	1	2	2E

3/40

FIG. 2B

1) 509 hits identified by Genome Threader only:

Ad2list	Redundant sequence display	BPD link	WWW link	Title	Organism	Div.	%ID	Query rqn.	Target rqn.	Aln. score	Conf.
↓	Red Seq. View	AAF71133.1 drill through Top50BlastHits	AAF71133.1	PRO2769	Homo sapiens	PRI	13.8% unmaskedSW	109-179	1-80	122	100% unmaskedGT
↓	Red Seq. View	CAB52192.1 drill through Top50BlastHits	CAB52192.1	G7c protein	Homo sapiens	PRI	9.6% unmaskedSW	7-117	20-124	82	99.18% unmaskedGT
↓	Red Seq. View	CAA82910.1 drill through Top50BlastHits	CAA82910.1	basic transcription factor 2, 44 kD subunit	Homo sapiens	PRI	11% unmaskedSW	5-164	61-225	75	99.18% unmaskedGT
↓	Red Seq. View	AAD21820.1 drill through Top50BlastHits	AAD21820.1	NG37	Homo sapiens	PRI	9.6% unmaskedSW	7-117	318-422	82	98.86% unmaskedGT
↓	Red Seq. View	BA220781.1 drill through Top50BlastHits	BA220781.1	Not given	Homo sapiens	PRI	16.4% unmaskedSW	5-114	1836-1950	78	98.52% unmaskedGT
↓	Red Seq. View	AAA36154.1 drill through Top50BlastHits	AAA36154.1	Not given	Homo sapiens	PRI	13.2% unmaskedSW	5-112	10-137	79	98.16% unmaskedGT
↓	Red Seq. View	AAF03046.1 drill through Top50BlastHits	AAF03046.1	candidate tumor suppressor protein DICE1	Homo sapiens	PRI	13.8% unmaskedSW	5-113	4-131	79	97.07% unmaskedGT
↑	Red Seq. View	AAC74654.1 drill through Top50BlastHits	AAC74654.1	orf, hypothetical protein	Escherichia coli	BCT	16.7% unmaskedSW	5-89	250-335	78	95.09% unmaskedGT
↓	Red Seq. View	AAB0942.1 drill through Top50BlastHits	AAB0942.1	breast cancer suppressor candidate 1	Homo sapiens	PRI	16.2% unmaskedSW	63-156	2-90	80	93.33% unmaskedGT
↓	Red Seq. View	AA67537.1 drill through Top50BlastHits	AA67537.1	glycoprotein IIIa	Homo sapiens	PRI	18.9% unmaskedSW	3-112	111-251	72	92.08% unmaskedGT reverse Hit
↓	Red Seq. View	AAA52589.1 drill through Top50BlastHits	AAA52589.1	Not given	Homo sapiens	PRI	19.6% unmaskedSW	3-112	138-277	74	91.34% unmaskedGT reverse Hit
↓	Red Seq. View	AAB71380.1 drill through Top50BlastHits	AAB71380.1	platelet membrane glycoprotein IIIa beta subunit	Homo sapiens	PRI	19.6% unmaskedSW	3-112	138-277	74	91.34% unmaskedGT reverse Hit
↓	Red Seq. View	AAA35927.1 drill through Top50BlastHits	AAA35927.1	Not given	Homo sapiens	PRI	19.6% unmaskedSW	3-112	138-277	74	91.34% unmaskedGT

File Edit View Go Communicator

Go to: <http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl>

630 out of these 645 PSI-BLAST matches were identified using 'positive iterations':

[illegible]

15 out of these 645 PSI-BLAST matches were identified using 'negative iterations':

BABYI1Z7T,CAN1206F,F3903WQ,T7E2E,OAF358H,QA1P77G,AADN2HL,AAPB7GL,AAJ33BL,J5A1UOI,ARCI9IL,JAA85SL,AAP6LWV,ARC5IRZ,AAC5294F,QAIUL.

A2) Genome Threader Matches:

27229 matches found by Genome Threading:

[illegible]

5/40

inpharmatica

Redundant Sequence Display

File Edit Search Type Help

Contains:
2 Sub-sequences.
0 PROSITE hits.
0 PRINTS hits.

Representative code: BAA15585.1 Length: 427 Organism name: *Escherichia coli* [...additional annotation]

Aligned sequences

Code
BAA15585.1
P76235
AAC74854.1

Primary database information

GenBank	protein_id	Details
BAA15585.1	Tax ID: 562	Organism name: <i>Escherichia coli</i>
	EC Number: Not available	Gene Name: yzcC
	Description: Not given	
	Other links:	
	GI: 1796412 TAXON: 562	

SWISS-PROT

Code	Details
P76235	Tax ID: 562
	Organism name: <i>Escherichia coli</i>
	Gene Name: YEAH
	EC Number: Not available
	Description: HYPOTHETICAL 49.4 KDA PROTEIN IN GAPA-RND INTERGENIC REGION.

Tools

Sequences

Lipids

Start

inpharmatica

BP04-FEI.4

inpharmatica

Fagan, Richa...

Inbox - Outb...

Target Min...

Microsoft Po...


FIG. 3

6/40
FIG. 4

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop


Bookmarks Location: <http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi>



Pfam

Protein families database of alignments and HMMs

Home | [Keyword search](#) | [Protein search](#) | [DNA search](#) | [Browse Pfam](#) | [Taxonomy search](#) | [Help](#)




Results for [gi|1788084|gb|AAC74854.1](#)

There were no matches to Pfam-A (including borderline matches) for [gi|1788084|gb|AAC74854.1](#)

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_39416	233	423	3.7e-103	Align


[427 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: [Hypertext linked to swisspfam](#) ☐

Query [gi|1788084|gb|AAC74854.1|233-423](#) matching [Pfam-B_39416](#)

```

YEAH_ECOLI 233 DLRYKNYEKRPDPSSQAVMFCMDVSGSMDQSTKDMAKRFYILLYLFLSR 282
gi|1788084|gb|AAC74854.1| 233 DLRYKNYEKRPDPSSQAVMFCMDVSGSMDQSTKDMAKRFYILLYLFLSR 282

YEAH_ECOLI 283 TYKNVEVVYIRNHTQAKEVDEMEFFYSQETGGTIVSSALKLMDEVVKERY 332
gi|1788084|gb|AAC74854.1| 283 TYKNVEVVYIRNHTQAKEVDEMEFFYSQETGGTIVSSALKLMDEVVKERY 332

YEAH_ECOLI 333 NPAQWNIYAAQASDGDNWADD SPLCHEILAKKLLPVVRYYSYIEITRRRH 382
gi|1788084|gb|AAC74854.1| 333 NPAQWNIYAAQASDGDNWADD SPLCHEILAKKLLPVVRYYSYIEITRRRH 382

YEAH_ECOLI 383 QTLWREYENLQSTFDNFMQHIRDQDDIYPVVFRELPHKQNA 423
gi|1788084|gb|AAC74854.1| 383 QTLWREYENLQSTFDNFMQHIRDQDDIYPVVFRELPHKQNA 423
  
```

Align to family

If you think there is anything wrong with this script, please contact [Pfam](#)

100%

FIG. 5

7/40


File Edit View Go Communicator Help	
Bookmarks Location: http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=	
LOCUS	AAC74854 427 aa BCT 01-DEC-2000
DEFINITION	orf, hypothetical protein [Escherichia coli K12].
ACCESSION	AAC74854
PID	g1788084
VERSION	AAC74854.1 GI:1788084
DBSOURCE	locus AE000273 accession AE000273.1
KEYWORDS	
SOURCE	Escherichia coli K12.
ORGANISM	Escherichia coli K12 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE	1 (residues 1 to 427)
AUTHORS	Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, M.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
TITLE	The complete genome sequence of Escherichia coli K-12
JOURNAL	Science 277 (5331), 1453-1474 (1997)
MEDLINE	97426617
PUBMED	9278503
REFERENCE	2 (residues 1 to 427)
AUTHORS	Blattner, F.R.
TITLE	Direct Submission
JOURNAL	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE	3 (residues 1 to 427)
AUTHORS	Blattner, F.R.
TITLE	Direct Submission
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE	4 (residues 1 to 427)
AUTHORS	Plunkett, G. III.
TITLE	Direct Submission
JOURNAL	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT	This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names. Method: conceptual translation.
FEATURES	Location/Qualifiers
source	1..427 /organism="Escherichia coli K12" /strain="K12" /sub_strain="MG1655" /db_xref="taxon:83333"
Protein	1..427 /function="orf; Unknown" /product="orf, hypothetical protein"
CDS	1..427 /gene="peaH" /coded_bp="1788078:6385..7668" /transl_table=11 /note="o427; This 427 aa ORF is 28 pct identical (43 gaps) to 327 residues of an approx. 312 aa protein YZDC_BAC5U SW: P45742"
ORIGIN	1 mtwfidrrln gknksmvnrq rfrprkaqi kqsiseaink rsvtdvdsge svstiptedis 61 epmfhgqrgg lhrvhpgnd hfvndrier pggggggsgs gggqasqdge gqdeivqis 121 kdeydlilfe dlapnlkqn qqrqltepkt hrsgptangv panisvvrsl qnslarrtam 181 taokrrrelha leenlaish sepaqlleee rlrkeiaelr akiervpfid tfdlryknye
100%	

8/40

FIG. 6A

File Edit Search Type: Help

Accession Code Query



Tools

- ←
-
- ?
- ?
- ?
- X

Sequences

- ? ?
- ? ?
- ? ?
- ? Acc y123

Ligands

- ? RDP NRD FRD
- ? Phe Ala Glu

Select a database and enter a code:

Database	
GenBank	
Accession	<input checked="" type="radio"/>
Accession Version	<input type="radio"/>
GI-NUC	<input type="radio"/>
protein_id	<input type="radio"/>
GI-NUC	<input type="radio"/>
SWISS-PROT	
Accession	<input type="radio"/>
ID	<input type="radio"/>
PDB	
ID	<input type="radio"/>

AAC74854.1

Submit Reset

9/40

File Edit Search Type Help

Aligned Sequence Display

Query details:
AAC74854.1:orf, hypothetical protein

Tools

Sequences

Ligands

Page 1 of 2

Total hits: 36

Selected: 0

Total selected: 0

View alignment

Select all

Deselect all

Inpharmatica Genome Threader results:

Cluster Detail	Code	Title	Organism	%ID	Query rgn.	Target rgn.	Aln. score	Method	Confidence
1LFAA	CD11A	I-DOMAIN WITH BOUND MN++	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(95%)
1ZOOA	CD11A	I-DOMAIN WITH BOUND MAGNESI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(95%)
1ZOOB	CD11A	I-DOMAIN WITH BOUND MAGNESI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(95%)
1ZOPB	CD11A	I-DOMAIN WITH BOUND MAGNESI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(95%)
1ZOPA	CD11A	I-DOMAIN WITH BOUND MAGNESI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(95%)
1BHO1	MAC-1	I-DOMAIN MAGNESIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low(94%)
1IDN1	MAC-1	I-DOMAIN METAL FREE	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low(94%)
1BHQ2	MAC-1	I-DOMAIN CADMIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low(94%)
1BHQ1	MAC-1	I-DOMAIN CADMIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low(94%)
1BHO2	MAC-1	I-DOMAIN MAGNESIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low(94%)
1IDN2	MAC-1	I-DOMAIN METAL FREE	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low(94%)
1ATZB	HUMAN VON WILLEBRAND FACTOR A3 DO...		<i>H.sapiens</i>	15.1	244 - 352	1 - 126	70	Local	Low(94%)
1ATZA	HUMAN VON WILLEBRAND FACTOR A3 DO...		<i>H.sapiens</i>	15.3	250 - 335	5 - 101	72	Local	Low(93%)
1DGQA	NMR SOLUTION STRUCTURE OF THE INSE...		<i>H.sapiens</i>	15.3	250 - 331	9 - 99	78	Local	Low(93%)
1IDO	I-DOMAIN FROM INTEGRIN CR3, MG2+ BO...		<i>H.sapiens</i>	15.1	250 - 365	4 - 128	75	Local	Low(92%)
1JLM	I-DOMAIN FROM INTEGRIN CR3, MN2+ BO...		<i>H.sapiens</i>	15.1	250 - 365	4 - 128	75	Local	Low(89%)
1ZON	CD11A	I-DOMAIN WITHOUT BOUND CATI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	72	Local	Low(83%)
1LFAA	CD11A	I-DOMAIN WITH BOUND MN++	<i>H.sapiens</i>	18.0	250 - 333	5 - 97	86	Local	Low(80%)
1OAKA	CRYSTAL STRUCTURE OF THE VON WILLE...		<i>H.sapiens</i>	20.3	251 - 301	1 - 59	72	Local	Low(72%)
1AUQ	A1-DOMAIN OF VON WILLEBRAND FACTOR		<i>H.sapiens</i>	21.2	250 - 301	10 - 68	89	Local	Low(66%)
1AO3A	A3-DOMAIN OF VON WILLEBRAND FACTOR		<i>H.sapiens</i>	15.2	245 - 331	2 - 99	63	Local	Low(53%)
1AO3B	A3-DOMAIN OF VON WILLEBRAND FACTOR		<i>H.sapiens</i>	15.2	245 - 331	2 - 99	63	Local	Low(53%)
1AOXB	I-DOMAIN FROM INTEGRIN ALPHA2-BETA1		<i>H.sapiens</i>	13.1	242 - 335	1 - 105	53	Local	Low(21%)
1RYPD	CRYSTAL STRUCTURE OF THE 20S PROTE...		<i>Saccharomyces</i>	13.1	246 - 305	166 - 226	58	Local	Low(21%)
1RYPB	CRYSTAL STRUCTURE OF THE 20S PROTE...		<i>Saccharomyces</i>	13.1	246 - 305	166 - 226	58	Local	Low(21%)
1CK4B	CRYSTAL STRUCTURE OF RAT A1B1 INTEG...		<i>Rattus norvegicus</i>	10.5	250 - 381	5 - 96	55	Local	Low(20%)

GenBank

SWISS-PROT

PDB

FIG. 6B

10/40

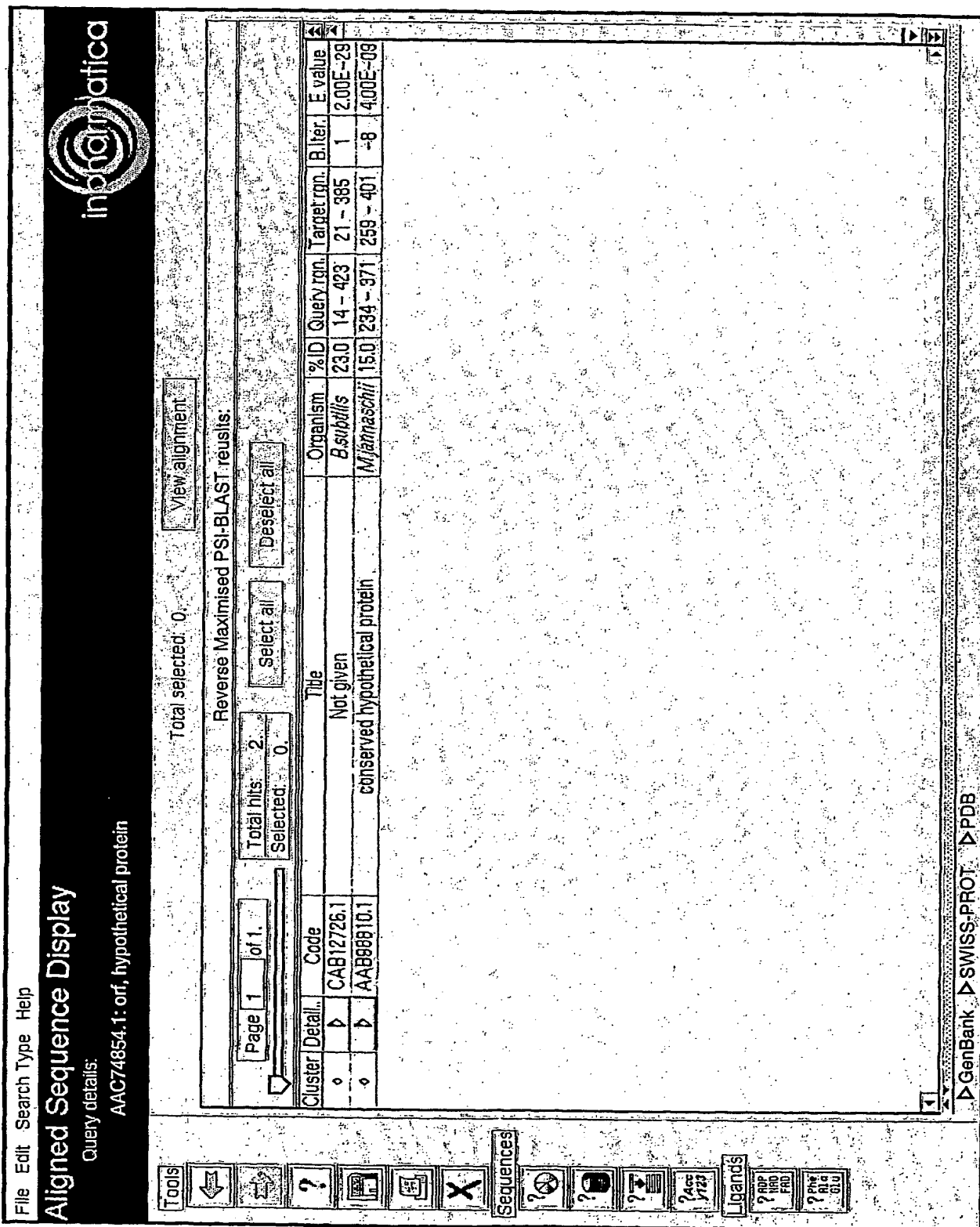


FIG. 6C

FIG. 7

AlEye output (January 4, 2002 3:07 PM)

11/40

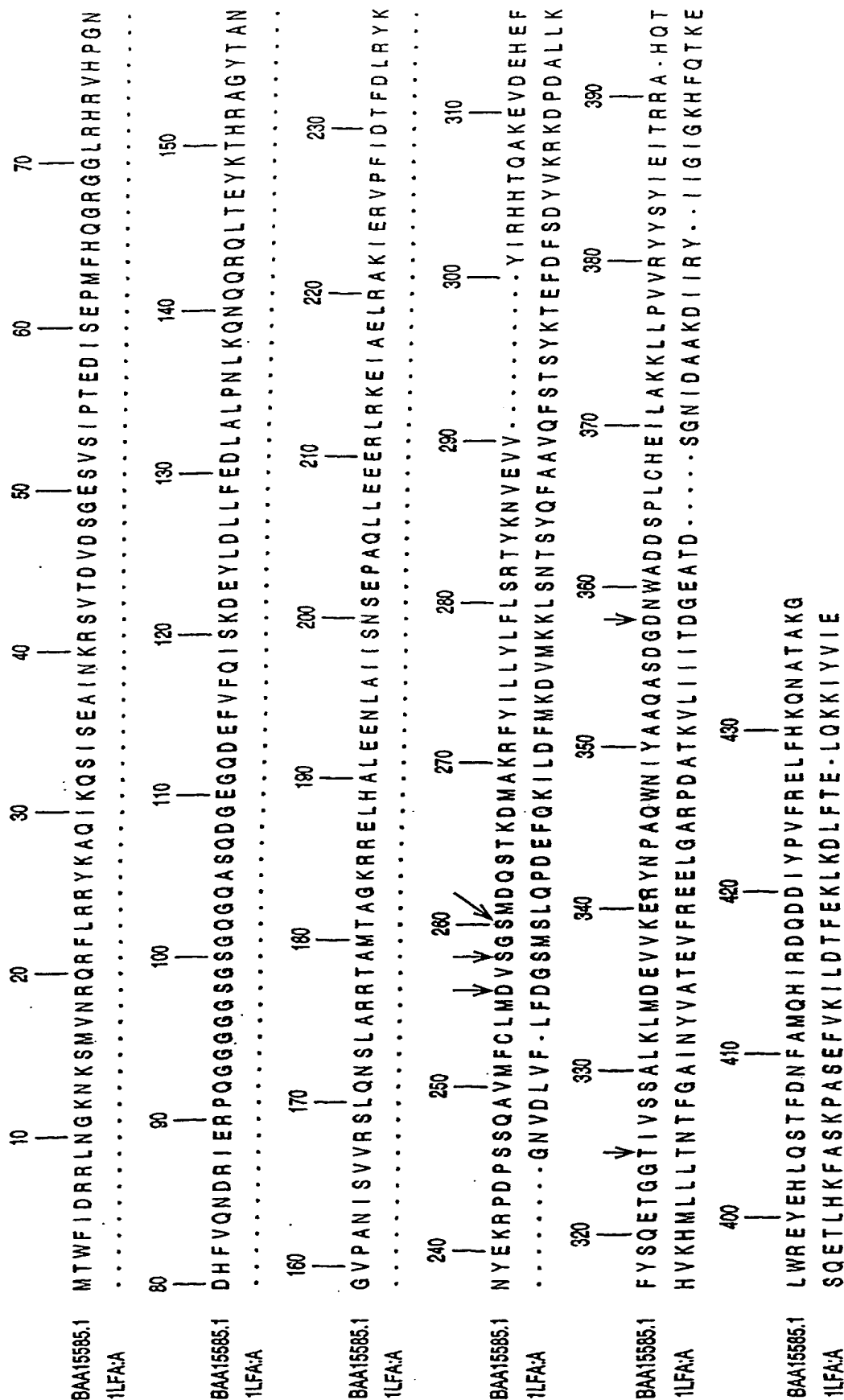


FIG. 8B

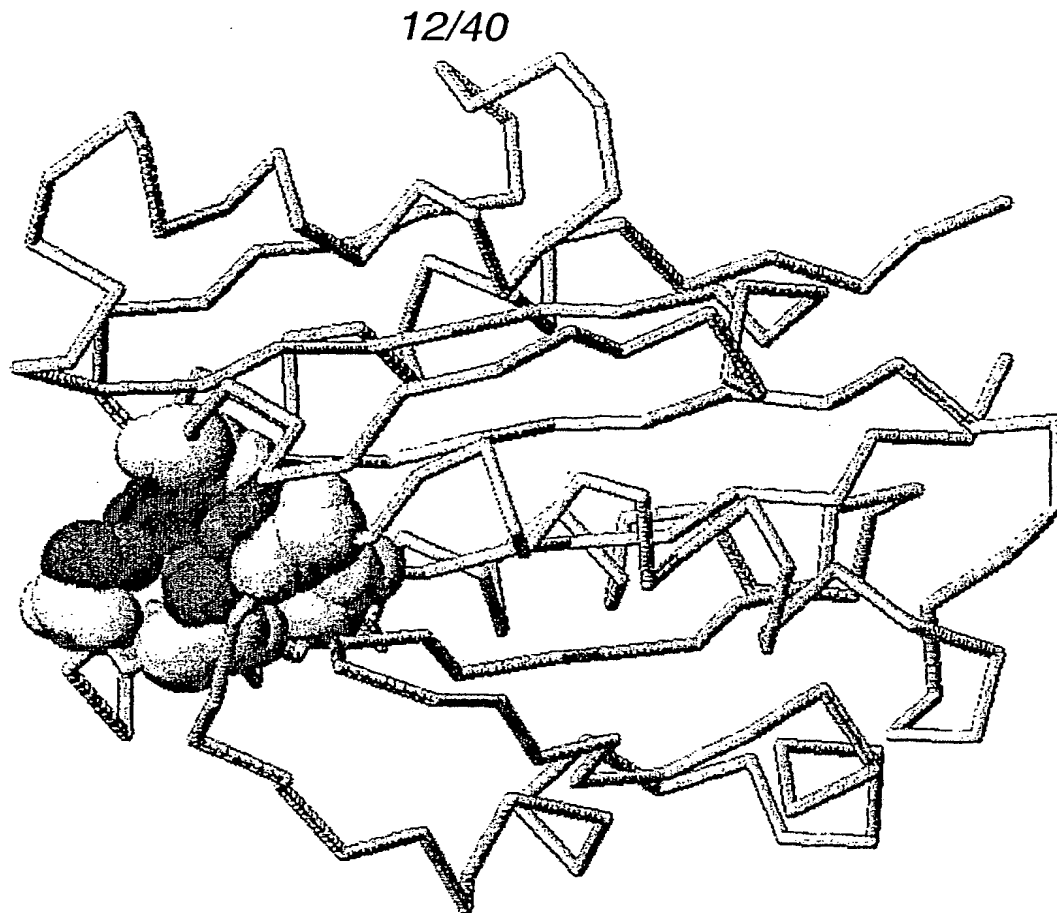
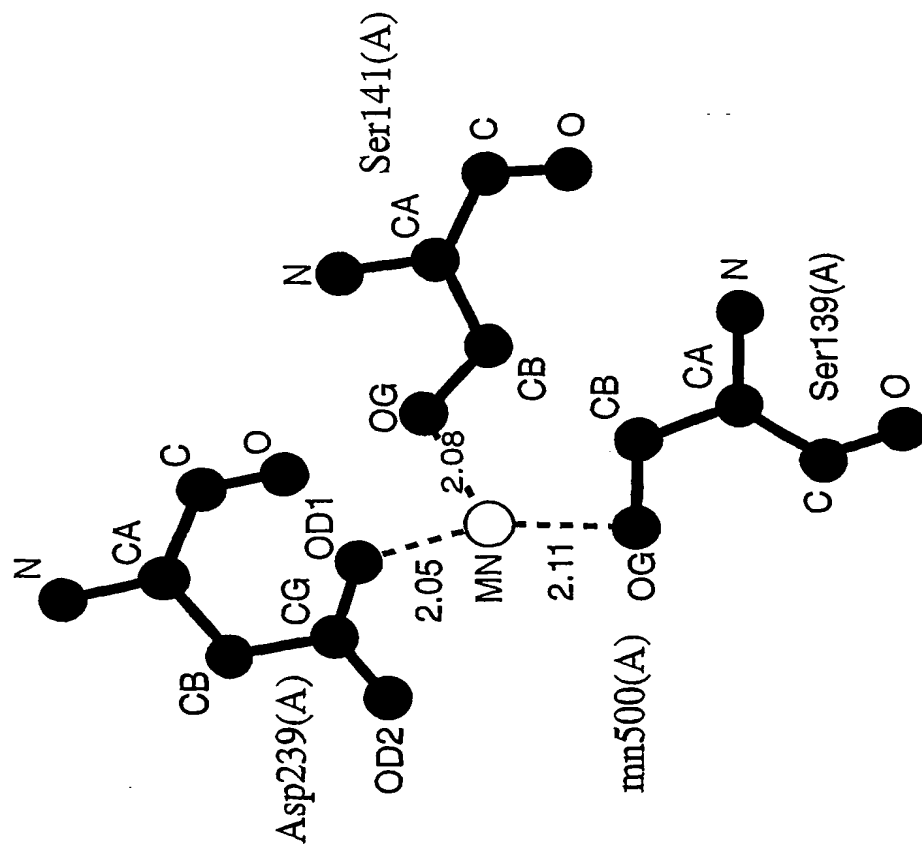


FIG. 8A

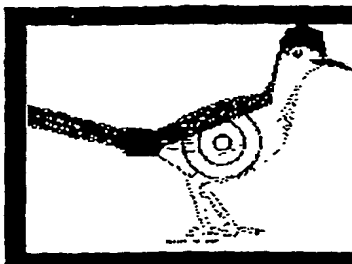


11fa: MN500 Chain [A]


FIG. 9

13/40

File Edit View Go Communicator Help
Bookmarks Location http://victoria.inpharmatica.co.uk/~volker/BPD3target.html



Target Mining Interface



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

14/40

FIG. 10A

File

Edit

View

Go

Communicator

Back

Forward

Reload

Home

Search

NetScope

Print

Security

Shop

Stop

Bookmarks

Location

http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/gat/targetBP03.pl#Mine2

2) 82 additional hits identified by both, Genome Threader and PSI-BLAST:

Credited Genome Threader and PSI-BLAST values are shown in red!

Accession	DPD info	VIMVEnt	Title	Organism	Div	%ID (GI/PS)	Query len (GI/PS)	Target len (GI/PS)	App. score (GI)	Conf. (GI)	1st iter. (PS)	Best iter. (PS)	Best E-value (PS)
U	F58193 cdt through Ips50BlastHit1 Red Sea View	E26192	INTEGRIN ALPHA-1 (LAMININ AND COLLAGEN RECEPTOR) (VLA-1) (CD49A)	Homo sapiens (Human)	PRI	51.8%, 51% unmaskedSW	2-200, 2-200	138-337, 139-337	487	100% unmaskedGI	1	2	2E-73
U	E17301 cdt through Ips50BlastHit1 Red Sea View	E17301	PLATELET MEMBRANE GLYCOPROTEIN IA PRECURSOR (GPIA) (COLLAGEN RECEPTOR) (INTEGRIN ALPHA-2) (VLA-2 ALPHA CHAIN) (CD49B)	Homo sapiens (Human)	PRI	99.5%, 100% unmaskedSW	2-200, 2-200	169-367, 169-367	471	100% unmaskedGI	1	1	1E-111
U	AA01258.1 cdt through Ips50BlastHit1 Red Sea View	AA01258.1	Integrin alpha-11 subunit precursor	Homo sapiens	PRI	45.2%, 45% unmaskedSW	2-200, 2-200	159-355, 159-355	469	100% unmaskedGI	1	2	4E-68
U	AA051918.2 cdt through Ips50BlastHit1 Red Sea View	AA051918.2	Integrin alpha 11 subunit precursor	Homo sapiens	PRI	45.2%, 45% unmaskedSW	2-200, 2-200	159-355, 159-355	469	100% unmaskedGI	1	2	4E-68
U	P39715 cdt through Ips50BlastHit1 Red Sea View	Q92715	COLLAGEN ALPHA (XII) CHAIN PRECURSOR.	Homo sapiens (Human)	PRI	28.1%, 31% unmaskedSW	2-186, 7-189	133-318, 2323-2485	437	100% unmaskedGI	1	2	2E-42
U	AA459491.1 cdt through Ips50BlastHit1 Red Sea View	AA459491.1	Not given	Homo sapiens	PRI	27.1%, 27% unmaskedSW	2-200, 2-200	144-338, 144-338	435	100% unmaskedGI	1	2	1E-60
U	AA459544.1 cdt through Ips50BlastHit1 Red Sea View	AA459544.1	Not given	Homo sapiens	PRI	27.1%, 27% unmaskedSW	2-200, 2-200	144-338, 144-338	435	100% unmaskedGI	1	2	1E-60
U	CA071722.1 cdt through Ips50BlastHit1 Red Sea View	CA071722.1	alpha2(BD)15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	26.9%, 26% unmaskedSW	5-195, 156-201	293-472, 1581-1626	424	100% unmaskedGI	2	3	5E-6
U	P21941 cdt through Ips50BlastHit1 Red Sea View	P21941	CARTILAGE MATRIX PROTEIN PRECURSOR (MATRILIN-1).	Homo sapiens (Human)	PRI	26.8%, 26% unmaskedSW	3-189, 3-198	271-455, 37-224	424	100% unmaskedGI	1	2	3E-46
U	AA01506.1 cdt through Ips50BlastHit1 Red Sea View	AA01506.1	type XII collagen	Homo sapiens	PRI	27.6%, 30% unmaskedSW	2-185, 2-195	133-318, 133-318	422	100% unmaskedGI	1	2	3E-50
U	AA031852.1 cdt through Ips50BlastHit1 Red Sea View	AA031852.1	Integrin subunit alpha 10 precursor	Homo sapiens	PRI	46.5%, 46% unmaskedSW	2-189, 2-199	162-359, 162-359	419	100% unmaskedGI	1	2	5E-63
U	CA003691.1 cdt through Ips50BlastHit1 Red Sea View	CA003691.1	matrin-4	Homo sapiens	PRI	26%, 28% unmaskedSW	1-189, 2-200	335-529, 28-223	411	100% unmaskedGI	1	2	1E-47

15/40

FIG. 10B

GT
Confidence

	Top50BlastHits Red. Seq. View	BA091707.1	Not given	Human sapiens	PRI	unmasked SW	unmasked GT	36-141	142	unmasked GT	3	5	2E-8
J	AAA36795.1 drill through Top50BlastHits	AAA36795.1	undulin 2	Homo sapiens	PRI	34.7%, 21% unmasked SW	100% unmasked GT	3-51, 37-93	121	100% unmasked GT	2	2	8E-6
J	Red. Seq. View												
J	AAD40367.1 drill through Top50BlastHits	AAD40367.1	calcium-activated chloride channel-2	Homo sapiens	PRI	15%, 15% unmasked SW	100% unmasked GT	312-421, 312-421	117	100% unmasked GT	5	5	1E-4
J	Red. Seq. View												
J	CAA67559.1 drill through Top50BlastHits	CAA67559.1	collagen VI- α -1 chain	Homo sapiens	PRI	21.1%, 21% unmasked SW	100% unmasked GT	7-92, 7-92	114	100% unmasked GT	3	3	5E-7
J	Red. Seq. View												
→	AAC76768.1 drill through Top50BlastHits	AAC76768.1	orf, hypothetical protein	Escherichia coli	BCT	13.2%, 11% unmasked SW	100% unmasked GT	267-384, 265-424	107	100% unmasked GT	-4	-5	3E-24
J	Red. Seq. View												
J	CAB43000.1 drill through Top50BlastHits	CAB43000.1	tellurite resistance	Escherichia coli	BCT	12.9%, 14% unmasked SW	100% unmasked GT	214-398, 208-380	102	100% unmasked GT	-6	-7	6E-22
J	Red. Seq. View												
J	AAA60114.1 drill through Top50BlastHits	AAA60114.1	platelet membrane glycoprotein IIb	Homo sapiens	PRI	14%, 13% unmasked SW	98.52% unmasked GT	145-187, 145-187	111	98.52% unmasked GT	-7	-8	8E-5
J	Red. Seq. View						reverse Hit						

17/40

inpharmatica

Redundant Sequence Display

Contains:
2 Sub-sequences.
0 PROSITE hits.
0 PRINTS hits.

File Edit Search Type Help

Representative code: P03818 Length: 427 Organism name: *Escherichia coli* [..additional annotation]

Aligned sequences

Code	Sequence
P03818	↓
AAC76768.1	
AAA62097.1	

Primary database information

SWISS-PROT	
Code	Details
P03818	Tax ID: 562
EC Number: Not available	Organism name: <i>Escherichia coli</i>
	Gene Name: y1em
Description: HYPOTHETICAL 49.6 KDA PROTEIN IN ASNA-KUP INTERGENIC REGION.	
Other links:	
ECOGENE: EG-H730 PIR: A04443	

GenBank	
protein_id	Details
AAC76768.1	Tax ID: 562
	Organism name: <i>Escherichia coli</i>
	EC Number: Not available
	Gene Name: y1em
Description: orf_hypothetical protein	

Tools

Sequences

Ligands

start BP04-F... inphar... Fagen... Inbox... NCBI S... Microso... 4:01 PM


FIG. 11

18/40

FIG. 12


File Edit View Go Communicator Help

Bookmarks Location: <http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi>



Pfam
Protein families database of alignments and HMMs

Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help

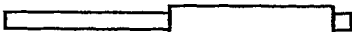


Results for gi|2367274|gb|AAC76768.1|

There were no matches to Pfam-A (including borderline matches) for gi|2367274|gb|AAC76768.1|

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B 15204	204	408	2.4e-108	Align


[427 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: ☒ Hypertext linked to swisspfam ☐

Query gi|2367274|gb|AAC76768.1|/204-408 matching Pfam-B 15204

```

YIEM_ECOLI 204 DILRLPPPELATLGITELLEYFYRLVEKQLLTYRLNGESWREKVIERP 253
gi|2367274|gb|AAC76768.1| 204 DILRLPPPELATLGITELLEYFYRLVEKQLLTYRLNGESWREKVIERP 253

YIEM_ECOLI 254 VHKDYDEQPRGPFIVCVDTSGSHG6FNEQCAKAPCLALMRIALAENRRCY 303
gi|2367274|gb|AAC76768.1| 254 VHKDYDEQPRGPFIVCVDTSGSHG6FNEQCAKAPCLALMRIALAENRRCY 303

YIEM_ECOLI 304 IMLFSTEIVRYELSGPQ6IEQAIRFLSQQFRGGTDLASCYRAIMERLQSR 353
gi|2367274|gb|AAC76768.1| 304 IMLFSTEIVRYELSGPQ6IEQAIRFLSQQFRGGTDLASCYRAIMERLQSR 353

YIEM_ECOLI 354 EWFADADAVVISDFIAQRLPDDVTSKVKELQRVHQHMFHAVMSAMGKPGI 403
gi|2367274|gb|AAC76768.1| 354 EWFADADAVVISDFIAQRLPDDVTSKVKELQRVHQHMFHAVMSAMGKPGI 403

YIEM_ECOLI 404 MRIFD 408
gi|2367274|gb|AAC76768.1| 404 MRIFD 408

```

Align to family

If you think there is anything wrong with this script, please contact [Pfam](#)

FIG. 13

19/40


File Edit View Go Communicator Help	
Bookmarks Location: http://www.ncbi.nlm.nih.gov/80/entrez/query.fcgi?cmd=Retrieve&db=protein	
LOCUS	ARC76768 427 aa BCT 01-DEC-2000
DEFINITION	orf, hypothetical protein [Escherichia coli K12].
ACCESSION	ARC76768
PID	g2367274
VERSION	ARC76768.1 GI:2367274
DBSOURCE	locus AE000451 accession AE000451.1
KEYWORDS	
SOURCE	Escherichia coli K12.
ORGANISM	Escherichia coli K12 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE	1 (residues 1 to 427)
AUTHORS	Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, R.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Maphew, S.F., Gregor, J., Davis, R.W., Kirkpatrick, M.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
TITLE	The complete genome sequence of Escherichia coli K-12
JOURNAL	Science 277 (5331): 1453-1474 (1997)
MEDLINE	97426617
PUBMED	9278503
REFERENCE	2 (residues 1 to 427)
AUTHORS	Blattner, F.R.
TITLE	Direct Submission
JOURNAL	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE	3 (residues 1 to 427)
AUTHORS	Blattner, F.R.
TITLE	Direct Submission
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE	4 (residues 1 to 427)
AUTHORS	Plunkett, G. III.
TITLE	Direct Submission
JOURNAL	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT	This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants M00301 and M01428 (from the Human Genome Project and HCHGR). The entire sequence was independently determined from E. coli K12 strain M01655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amherst.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with cc site nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
FEATURES	Method: conceptual translation.
source	Location/Qualifiers 1..427 /organism="Escherichia coli K12" /strain="K12" /sub_strain="M01655" /db_xref="taxon:83333"
Protein	1..427 /function="orf; Unknown" /product="orf, hypothetical protein"
CDS	1..427 /gene="pieM" /coded_by="complement(2367272:5249..6532)" /transl_table=11 /note="f427; sequence change joins ORFs pieD and pieM from earlier version"
ORIGIN	1 mrsrlkdarv ppelteevmc yqqsqllstp qfiavqlpqil dlhlrlnsyw aeqarglvda 61 nstitsalht lflqrwrsls lvgattingq lleeereqll sevqermals gqlepiladm 121 ntaagrlwda sagqlkrgdy qlivkygefi neqpelkrla eqqlgrsreak siprndaqae 181 tfrtmvrepv tpeqvvgllq qsdilrlly belatlgite leueferrly ehalltrrh

20/40

FIG. 14A

File Edit Search Type Help

Accession Code Query



Tools

- ←
-
- ?
- ?
- ?
- X

Sequences

- ? ?
- ? ?
- ? ?
- ? Acc y123

Ligands

- ? RDP NAD FAD
- ? Phe R1a Glu

Select a database and enter a code:

Database	
GenBank	
Accession	<input type="radio"/>
Accession Version	<input type="radio"/>
GI-ND	<input type="radio"/>
protein_id	<input type="radio"/>
GI-PID	<input type="radio"/>
SWISS-PROT	
Accession	<input type="radio"/>
ID	<input type="radio"/>
PDB	
ID	<input type="radio"/>

AAC76768.1

Submit Reset

FIG. 14B

File Edit Search Type Help

Aligned Sequence Display

Query details:
AAC76768.1:orf, hypothetical protein

Tools

Sequences

Ligands

Page 1 of 3

Total hits: 88

Selected: 0

Total selected: 0

View alignment

Select all

Deselect all

Inpharmatica Genome Threader results

Cluster	Detail	Code	Title	Organism	%ID	Query rgn	Target rgn	Aln. score	Method	Confidence
•	▶	1CK4B	CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN CR3, MG2+ BO...	<i>Rhovegicus</i>	12.6	267 - 398	6 - 144	158	Local	Certain(100%)
•	▶	1IDO	I-DOMAIN FROM INTEGRIN CR3, MG2+ BO...	<i>Hsaplens</i>	19.8	267 - 384	5 - 130	157	Local	Certain(100%)
•	▶	1JLM	I-DOMAIN FROM INTEGRIN CR3, MN2+ BO...	<i>Hsaplens</i>	19.8	267 - 384	5 - 130	156	Local	Certain(100%)
•	▶	1CK4A	CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN CR3, MG2+ BO...	<i>Rhovegicus</i>	12.6	267 - 398	6 - 144	149	Local	Certain(100%)
•	▶	1BHO1	MAC-1 DOMAIN MAGNESIUM COMPLEX	<i>Hsaplens</i>	19.8	267 - 384	5 - 130	146	Local	Certain(100%)
•	▶	1BHQ2	MAC-1 DOMAIN MAGNESIUM COMPLEX	<i>Hsaplens</i>	19.8	267 - 384	5 - 130	146	Local	Certain(100%)
•	▶	1BHQ1	MAC-1 DOMAIN MAGNESIUM COMPLEX	<i>Hsaplens</i>	19.8	267 - 384	5 - 130	146	Local	Certain(100%)
•	▶	1IDN2	MAC-1 DOMAIN METAL FREE	<i>Hsaplens</i>	19.8	267 - 384	5 - 130	146	Local	Certain(100%)
•	▶	1IDN1	MAC-1 DOMAIN METAL FREE	<i>Hsaplens</i>	19.8	267 - 384	5 - 130	146	Local	Certain(100%)
•	▶	1BHO2	MAC-1 DOMAIN MAGNESIUM COMPLEX	<i>Hsaplens</i>	19.8	267 - 384	5 - 130	146	Local	Certain(100%)
•	▶	1LFA8	CD11A I-DOMAIN WITH BOUND MN++	<i>Hsaplens</i>	15.9	264 - 381	1 - 124	143	Local	Certain(100%)
•	▶	1ZON	CD11A I-DOMAIN WITHOUT BOUND CAT++	<i>Hsaplens</i>	15.9	264 - 381	1 - 124	143	Local	Certain(100%)
•	▶	1LFAA	CD11A I-DOMAIN WITH BOUND MN++	<i>Hsaplens</i>	14.6	264 - 354	1 - 96	141	Local	Certain(100%)
•	▶	1ZOOA	CD11A I-DOMAIN WITH BOUND MAGNESI...	<i>Hsaplens</i>	14.6	264 - 354	1 - 96	141	Local	Certain(100%)
•	▶	1ZOPB	CD11A I-DOMAIN WITH BOUND MAGNESI...	<i>Hsaplens</i>	14.6	264 - 354	1 - 96	141	Local	Certain(100%)
•	▶	1ZOPA	CD11A I-DOMAIN WITH BOUND MAGNESI...	<i>Hsaplens</i>	14.6	264 - 354	1 - 96	141	Local	Certain(100%)
•	▶	1ZOOB	CD11A I-DOMAIN WITH BOUND MAGNESI...	<i>Hsaplens</i>	14.6	264 - 354	1 - 96	141	Local	Certain(100%)
•	▶	1DGOA	NMR SOLUTION STRUCTURE OF THE INSER...	<i>Hsaplens</i>	14.3	267 - 354	10 - 100	139	Local	Certain(100%)
•	▶	1ATZA	HUMAN VON WILLEBRAND FACTOR A3 DO...	<i>Hsaplens</i>	11.3	267 - 407	6 - 152	135	Local	Certain(100%)
•	▶	1OAKA	CRYSTAL STRUCTURE OF THE VON WILLE...	<i>Hsaplens</i>	11.3	267 - 398	10 - 150	131	Local	Certain(100%)
•	▶	1AOXA	I-DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>Hsaplens</i>	13.2	267 - 384	9 - 134	107	Local	Certain(100%)
•	▶	1GGSB	I-DOMAIN FROM INTEGRIN ALPHA1-BETA1	<i>N/A</i>	11.9	267 - 398	5 - 143	106	Local	Certain(100%)
•	▶	1GCSA	I-DOMAIN FROM INTEGRIN ALPHA1-BETA1	<i>N/A</i>	11.9	267 - 398	7 - 145	102	Local	Certain(100%)
•	▶	1AOXB	I-DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>Hsaplens</i>	13.3	267 - 398	11 - 149	98	Local	Certain(100%)
•	▶	1AOSA	A3 DOMAIN OF VON WILLEBRAND FACTOR	<i>Hsaplens</i>	10.7	267 - 398	8 - 145	98	Local	Certain(100%)
•	▶	1AOSB	A3 DOMAIN OF VON WILLEBRAND FACTOR	<i>Hsaplens</i>	10.7	267 - 398	8 - 145	98	Local	Certain(100%)
•	▶	1AUQ	A1 DOMAIN OF VON WILLEBRAND FACTOR	<i>Hsaplens</i>	11.8	250 - 398	4 - 158	97	Local	Certain(100%)
•	▶	1ATZB	HUMAN VON WILLEBRAND FACTOR A3 DO...	<i>Hsaplens</i>	10.7	267 - 398	9 - 146	96	Local	Certain(100%)
•	▶	1AGJB	NITROGEN REGULATORY BACTERIAL PRO...	<i>Ecol</i>	18.6	306 - 348	13 - 57	60	Local	Low(42%)
•	▶	1BKDS	COMPLEX OF HUMAN H-RAS WITH HUMA...	<i>Hsaplens</i>	14.4	141 - 292	100 - 202	59	Local	Low(34%)

GenBank

SWISS-PROT

PDB

File Edit Search Type Help

Aligned Sequence Display

Query details:

AAC76768.1: orf, hypothetical protein



Tools



View alignment

Total selected: 0

Reverse Maximised PSI-BLAST results:

Select all

Deselect all

Total hits: 508

Selected: 0

Page 1 of 17

Cluster Details

Code

Title

Organism

%ID

Query rgn

Target rgn

Bitter

E value

Cluster

Accession

Description

Accession

Description

Accession

Description

Accession

Description

Accession

Description

Accession

Description

Accession

Description

Accession

Description

Accession

Description

Accession

Description

Accession

Description

Accession

Description

Accession

Description

23/40

FIG. 15

AlEye output (December 13, 2000 3:07 PM)

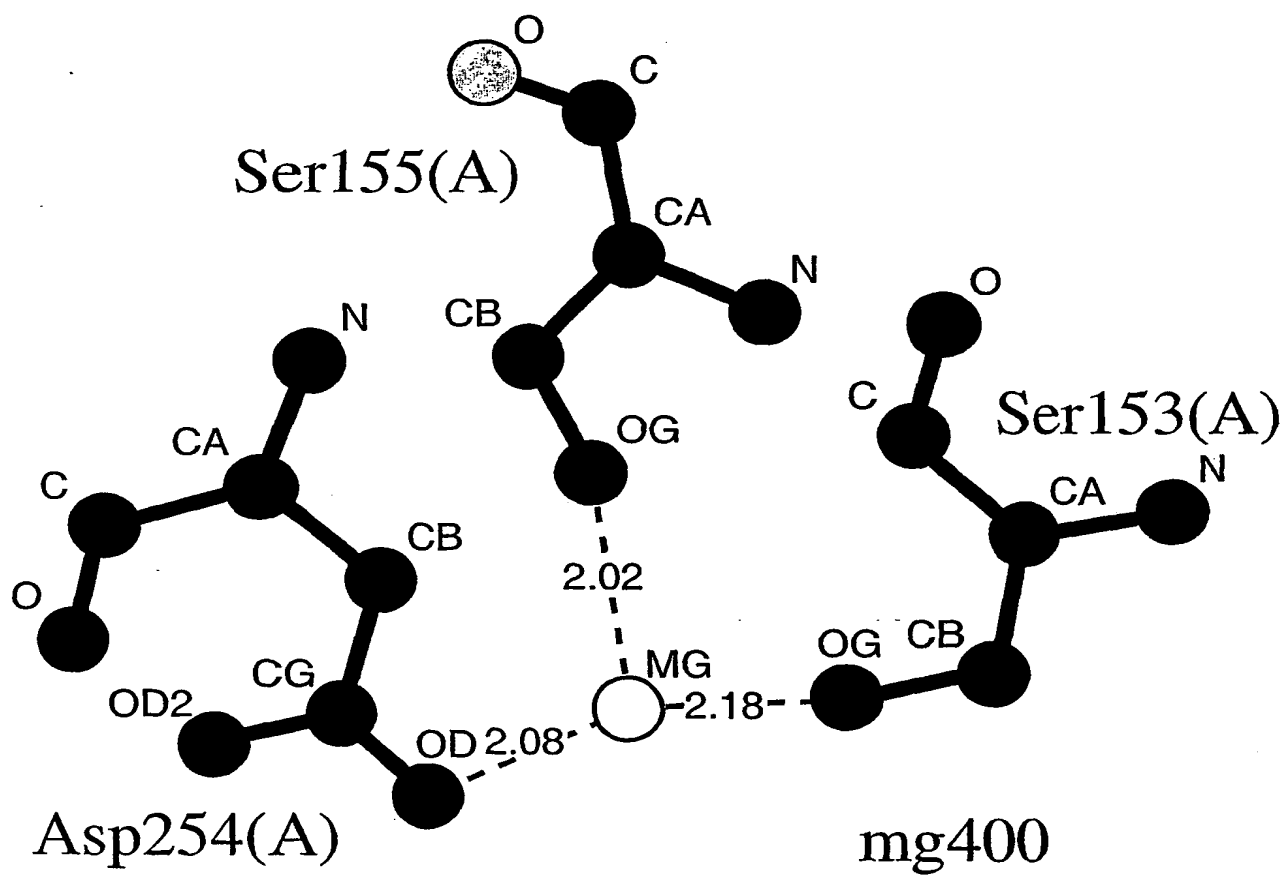
```

AAC76768.1 10 20 30 40 50 60 70
1A0X:A MRSRLKDAHVPPPELTTEVNCYQQSLLSTPQFIVQLPQILDLLHRLNSPWA EQARQLVDANSTITSA LHTLFLQRWR
-----
AAC76768.1 80 90 100 110 120 130 140 150
1A0X:A LSLIVQATTLNQQLLEEREQLLSEVQERMTLSGQLEPILADNNTAAAGRLWDM SAGQLKRGDYQLIVKYGEFLNEQP
-----
AAC76768.1 160 170 180 190 200 210 220 230
1A0X:A ELKRLAEQLGRSREAKSIPRNDAAQMETFRMTMVRPATVPEQVQGLQQSDDILRLPPELATLGIT ELEYEFYRRLVE
-----
AAC76768.1 240 250 260 270 280 290 300
1A0X:A KQLLTYYRLHGESWREKVIERPVVVHKDYDEQPRGPFIVCVDTSGSMGGFNEQCAKAFCLALMRI--ALAENRRCYIML
-----
AAC76768.1 310 320 330 340 350 360 370 380
1A0X:A -----SCPSLID-----VYVVCDESNSIYPWDA--VKNFLEKFVQGLDIGPTKTQVGLIQ
-----
AAC76768.1 390 400 410 420 430 440 450 460 470
1A0X:A FSTEI-VRYELSGPQG---IEQAIRFLSQQFRGGTDLASCFAIMERLQS-----REWFDADAVVISDFIAQRLPDD
YANNPRVVFNLTYYKTKEEMIVATSQTSQYGGDLTNTFGA IQYARKYAYSAASGGRRSATKYMVVVYTDGESH-DGSM
-----
AAC76768.1 480
1A0X:A VTSKVKELQRVHQHFFHAVAMSAHGKPGIMRIFDHIWRFDTGMRSRLLRRWR
LKAVIDQCNHDN-ILRFGIAYLGY-----LNRNALDTKNLIK EIKAIASIPTERYFFNVSD EAA LLEKAG
-----
AAC76768.1
1A0X:A TLGEQIFSI EGGT

```

24/40

FIG. 16A



1aox: MG400

25/40

FIG. 16B

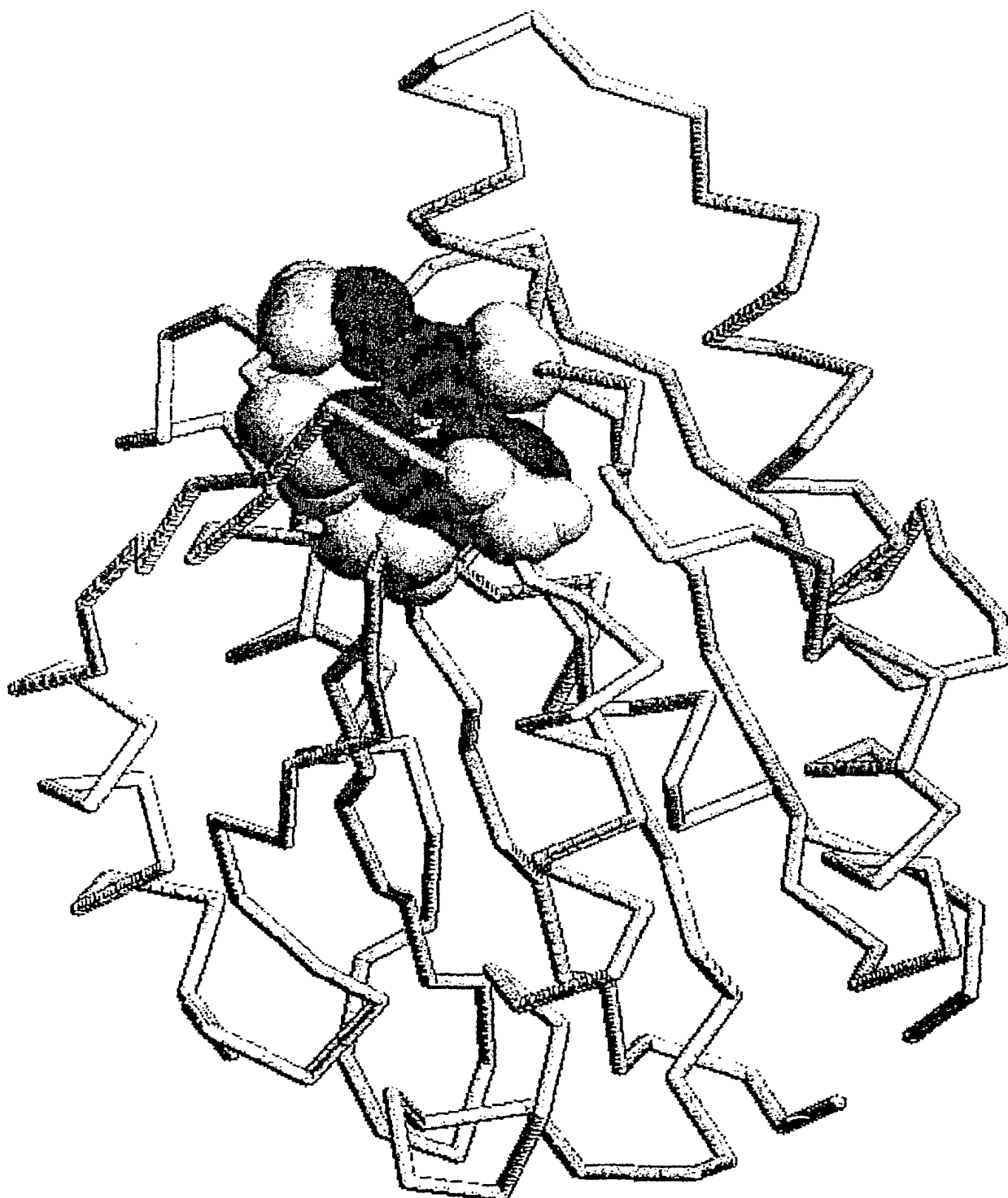
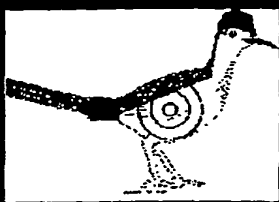



FIG. 17

26/40

File Edit View Go Communicator Help
Bookmarks Location: http://victoria.inpharmatica.co.uk/~volker/BPD3target.html

**Target Mining Interface**



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):
- OR
- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:
☐ Homo sapiens ☐ Rattus norvegicus (Rat) ☐ Mus musculus (Mouse) ☐ Danio rerio (Zebra fish)

100%

FIG. 18A

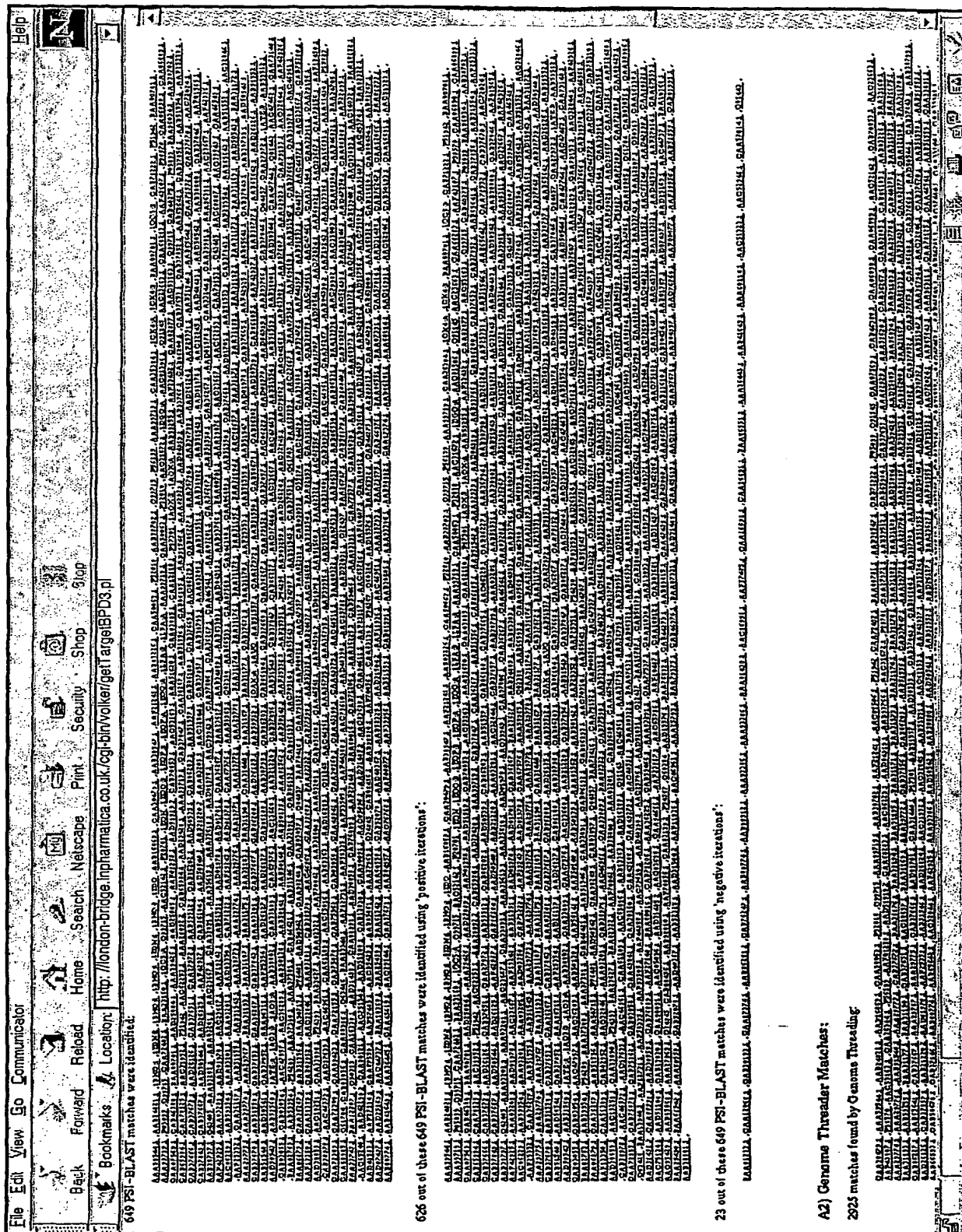
27/40

File Edit View Go Communicator Help											
Back Forward Reload Home Search Netscape Print Security Stop											
Bookmarks Location: [http://London-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl]											
2) 81 additional hits identified by both, Genome Threader and PSI-BLAST:											
Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!											
Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st Iter. (PSI)
J	AAA59544.1 drill through Top50BlastHits Red Seq View	AAA59544.1	Not given	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
→	AAB24821.1 drill through Top50BlastHits Red Seq View	AAB24821.1	leukocyte integrin alpha chain	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
J	Q99715 drill through Top50BlastHits Red Seq View	Q99715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human)	PRI	28.9%, 26% unmaskedSW	2-186, 2-179	439-617, 2322-2494	456	100% unmaskedGT	1
J	AAB38702.1 drill through Top50BlastHits Red Seq View	AAB38702.1	cartilage matrix protein	Homo sapiens	PRI	28.9%, 25% unmaskedSW	2-186, 2-186	274-452, 40-221	446	100% unmaskedGT	1
J	AAC01506.1 drill through Top50BlastHits Red Seq View	AAC01506.1	type XII collagen	Homo sapiens	PRI	28.4%, 28% unmaskedSW	2-186, 2-186	137-318, 137-318	445	100% unmaskedGT	1
J	CAA72402.1 drill through Top50BlastHits Red Seq View	CAA72402.1	collagen type XIV	Homo sapiens	PRI	28.7%, 30% unmaskedSW	2-186, 2-186	6-185, 6-185	442	100% unmaskedGT	1
J	AAB38547.1 drill through Top50BlastHits Red Seq View	AAB38547.1	leukocyte integrin alpha d chain	Homo sapiens	PRI	61%, 60% unmaskedSW	1-187, 1-187	148-334, 148-334	439	100% unmaskedGT	1
J	CAB71222.1 drill through Top50BlastHits Red Seq View	CAB71222.1	dJ238D15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	27.1%, 22% unmaskedSW	1-186, 2-186	293-472, 1430-1620	439	100% unmaskedGT	1
J	CAA07569.1 drill through Top50BlastHits Red Seq View	CAA07569.1	matrin-4	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-187	344-525, 34-215	418	100% unmaskedGT	2
J	CAB46380.1 drill through Top50BlastHits Red Seq View	CAB46380.1	dJ453C12.3 (matrin-4)	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-186	385-566, 34-214	418	100% unmaskedGT	2

28/40

FIG. 18B

File Edit View Go Communicator	Back Forward Reload Home Search Neilscape Print Security Shop Stop	Bookmarks Location: http://london-bridge.inpharmatica.co.uk/cgi-bin/lookup/getTargetBPD3.pl	Help
Red Sea View	AA018111	60 kDa RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO) (RORNP) (SUOREN SYNDROME TYPE A ANTIGEN (88-A))	66
Red Sea View	CAB52192.1	67c protein	76
Red Sea View	BAA75899.1	N-copine	73
Red Sea View	P28012	INTEGRIN BETA-8 PRECURSOR	78
Red Sea View	AA021820.1	NC37	78
Red Sea View	CA465725.1	Sec23 protein	74
Red Sea View	AA015820.1	copine 1	67
Red Sea View	AA063280.1	polymerase	55
Red Sea View	CA887610.1	d1308K20.3 (Copine 1 (similar to KIAA0556))	67
Red Sea View	AA055532.1	Not given	56
Red Sea View	P10155	60 KDA RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO) (RORNP) (SUOREN SYNDROME TYPE A ANTIGEN (88-A))	56
Red Sea View	AA015863.1	hexokinase 1	75
Red Sea View	BAA92872.1	KIAA1434 protein	67
Red Sea View	CAA10335.1	Sec24B protein	61
Red Sea View	AA043766.1	protocadherin gamma A11 short form protein	66
Red Sea View	AA043714.1	protocadherin gamma A11	66
Red Sea View	CA408533.1	annexin 31 (annexin XXXI)	58
Red Sea View	AA052846.1	hexokinase 1	73
Red Sea View	AA058185.1	Integrin beta-7 subunit	67
Red Sea View	AA055421.1	Integrin variant beta4E	73



30/40

FIG. 19

Biopendium additional annotation page - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: /london-bridge/inpharmatica/BPDEV1/cgi-bin/isp.pl?rep_esn=55817&password=caiss_app&username=caiss_app&role=sid=BPDEV1 What's Related

Aligned annotation view for P10155 (downloading image...)

1: A4435603.1
Rep: P10155

7 50 100 150 200 250 300 350 400 450 500

Primary database information: ☒ **Prosite matches** ☒ Prints matches

Secondary database information: ☒ **Prosite matches** ☒ Prints matches

Inpharmatica calculated information: ☒ **Prosite matches** ☒ Prints matches

Sequence information

Source databank	SWISSPROT
Accession code	P10155
Gene name	SSA2 OR R060
Download sequence in FASTA format	

Links to other resources:

View custom hyperlinks

Search in ExPASy

European Bioinformatics Institute	PR000504
Online Mendelian Inheritance in Man	234700
Online Mendelian Inheritance in Man	60063
Protein Information Resource	A30596
Protein Information Resource	A31760

Swissprot comments:

FUNCTION: UNKNOWN

FOUR SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT EACH OF


Document Done

31/40
FIG. 20

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop


Bookmarks Location: <http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi>



Pfam

Protein families database of alignments and HMMs

Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help



Results for gi|133251|sp|P10155|RO60_HUMAN

There were no matches to Pfam-A (including borderline matches) for gi|133251|sp|P10155|RO60_HUMAN

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_8344	1	194	2.3e-103	Align
Pfam-B_10162	195	538	1.8e-165	Align

[538 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: ☐ [Hypertext linked to swisspfam](#)

Query gi|133251|sp|P10155|RO60_HUMAN/1-194 matching Pfam-B_8344

```

092787 1 MEE SVNQM QPLNEKQIANSDG YVWQVTD MNR LHRFLCF GSEG GTTYIKE 50
gi|133251|sp|P10155|RO60_HUMAN 1 MEE SVNQM QPLNEKQIANSDG YVWQVTD MNR LHRFLCF GSEG GTTYIKE 50
092787 51 QKL GLENAEALIRLIEDGRGCEVIQEIKSF SQEGRTTKQEPMLFALAI CS 100
gi|133251|sp|P10155|RO60_HUMAN 51 QKL GLENAEALIRLIEDGRGCEVIQEIKSF SQEGRTTKQEPMLFALAI CS 100
092787 101 QCS DI STKQAAF KAV SEV CRIP THLFTFIQFKKDLKESMKCGMWRALRK 150
gi|133251|sp|P10155|RO60_HUMAN 101 QCS DI STKQAAF KAV SEV CRIP THLFTFIQFKKDLKESMKCGMWRALRK 150
092787 151 AIADWYNEKGGMALALAVTKYKQRNGWSHKDLLRLSHLKP SSEG 194
gi|133251|sp|P10155|RO60_HUMAN 151 AIADWYNEKGGMALALAVTKYKQRNGWSHKDLLRLSHLKP SSEG 194
  
```

[Align to family](#)

Query gi|133251|sp|P10155|RO60_HUMAN/195-538 matching Pfam-B_10162

```

008848 195 LAIVTKYITKGWKEVMEETKEKALSVREKLLKYLEAVEKVKRTKDDLEV 244
gi|133251|sp|P10155|RO60_HUMAN 195 LAIVTKYITKGWKEVMEETKEKALSVREKLLKYLEAVEKVKRTKDDLEV 244
008848 245 IMLEENQVLUREHLTHNLKSKEVWKALLQEMPLTALLRNLGKMTANSVL 294
THLTERH+LUREHLTHNLKSKEVWKALLQEMPLTALLRNLGKMTANSVL
008848 295 EPGNSEVSLICEKLSHEKLLKARIHPFHVLI ALETYTRAGHLRGLKWKY 344
gi|133251|sp|P10155|RO60_HUMAN 295 EPGNSEVSLICEKLSHEKLLKARIHPFHVLI ALETYTRAGHLRGLKWKY 344
008848 345 PDKDILQALDRAFYTTFTVETPT6KRFLAVDV SASMNQRALGSVLHAST 394
  
```

32/40
FIG. 21

File Edit View Go Communicator		Help
Back Forward Reload Home Search Netscape Print Security Stop		
Bookmarks Location: http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=Prot		
LOCUS	R060_HUMAN 538 aa PRI 01-FEB-1996	
DEFINITION	60 KD RO PROTEIN (60 KD RIBONUCLEOPROTEIN RO) (RORNP) (SJOGREN SYNDROME TYPE A ANTIGEN (SS-A)).	
ACCESSION	P10155	
PID	g133251	
VERSION	P10155 GI:133251	
DBSOURCE	swissprot: locus R060_HUMAN, accession P10155. class: standard. created: Mar 1, 1989. sequence updated: Mar 1, 1989. annotation updated: Feb 1, 1996. xrefs: gi: gi: 177782, gi: gi: 177783, gi: gi: 387656, gi: gi: 387657, gi: gi: 86722, gi: gi: 107626 xrefs (non-sequence databases): MIM 600053, MIM 234700, PROSITE P500030	
KEYWORDS	Ribonucleoprotein; RNA-binding; Systemic lupus erythematosus; Antigen.	
SOURCE	human.	
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (residues 1 to 538)	
AUTHORS	Deutscher, S.L., Marley, J.B. and Keene, J.D.	
TITLE	Molecular analysis of the 60-kDa human Ro ribonucleoprotein	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 85 (24), 9479-9483 (1988)	
MEDLINE	89071722	
REMARK	SEQUENCE FROM N.A	
REFERENCE	2 (residues 1 to 538)	
AUTHORS	Ben-Chetrit, E., Gandy, B.J., Tan, E.M. and Sullivan, K.F.	
TITLE	Isolation and characterization of a cDNA clone encoding the 60-kD component of the human SS-A/Ro ribonucleoprotein autoantigen	
JOURNAL	J. Clin. Invest. 83 (4), 1284-1292 (1989)	
MEDLINE	89198084	
REMARK	SEQUENCE FROM N.A.	
COMMENT	<p>This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. The original entry is available from http://www.expasy.ch/sprot and http://www.ebi.ac.uk/sprot</p> <p>[FUNCTION] UNKNOWN.</p> <p>[SUBUNIT] RO SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT, EACH OF WHICH IS COMPLEXED WITH A 60 KD PROTEIN. RO RNPS MAY ALSO CONTAIN AN ADDITIONAL 52 KD PROTEIN.</p> <p>[SUBCELLULAR LOCATION] CYTOPLASMIC.</p> <p>[DISEASE] SERA FROM PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS OFTEN CONTAIN ANTIBODIES THAT REACT WITH THE NORMAL CELLULAR RO PROTEIN AS IF THESE ANTIGEN WAS FOREIGN.</p> <p>[SIMILARITY] CONTAINS 1 RNA RECOGNITION MOTIF (RNP).</p> <p>[SIMILARITY] STRONG, TO XENOPUS 60 KD RO PROTEIN.</p>	
FEATURES	Location/Qualifiers source 1..538 /organism="Homo sapiens" /db_xref="taxon:9606" Protein 1..538 /product="60 KD RO PROTEIN" Region 93..98 /region_name="Domain" /note="RNA-BINDING (RNP2) (BY SIMILARITY)." Region 124..131 /region_name="Domain" /note="RNA-BINDING (RNP1) (BY SIMILARITY)." Region 239 /region_name="Conflict" /note="K -> R (IN REF. 2)." Region 515..538 /region_name="Conflict" /note="GMLDMCGFDTCALDVIRNFTLDMI -> ALQNTLLNKSF (IN REF. 2)." 	
ORIGIN	1 meesvnmqgp lnekqiansq dgyvwqvtdm nrhrflcsg seggtypike qklglенаа 61 lirliedrgy cevigeiksf sgegrttkqe pmlfalaics qosdistkqa afkavsever 121 ipthlftfiq fkkdkesmk egmwgralk aiadwynekq gmalalavtk pzqngwshk 181 dlrlrlshlkp sseglaiavtk yitkgwkevh elykekalsv etekllkyle avekvkrtd 241 elevihliee hrlvrehllt nhlkskevkk allqempita llnlglkmta nsvlepgnse 301 vslvceklen eklkkarih pphilialet pktghglrgk lkwrpdeeil kaldaafpht 361 fktveptgkr fllavdvsas mnqrvglsil nastvaaame mvvtrtekds yvvaifdemv 421 pcpvttamt1 qqvlmamsqi paggtdeclp miwaqktntp advfivftdn etfaggvghpa 481 ialreyrkkm dipaklivcg mtsngftiad pddrgmldmc gfdtgaldvi rnftldmi 	

33/40

FIG. 22A

File Edit Search Type Help

Accession Code Query

informatics

Tools

Sequences

Ugands

Select a database and enter a code:

Database	
GenBank	
Accession	<input type="radio"/>
Accession Version	<input type="radio"/>
GI-NUD	<input type="radio"/>
protein Id	<input type="radio"/>
GI-PID	<input type="radio"/>
SWISS-PROT	
Accession	<input checked="" type="radio"/>
ID	<input type="radio"/>
PDB	
ID	<input type="radio"/>

P10155

Submit Reset

File Edit Search Type Help

FIG. 22B

Aligned Sequence Display

Query details:

P10155: 60 KDA RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO) (FORNP) (SJORGEN SYNDROME TYPE A ANTIGEN (SS-A)).



Tools: Sequences: Ligands:

Total selected: 1. View alignment

Inpharmatica Genome Threader results:

Page 1 of 2. Total hits: 31 Selected: 1.

Cluster Details	Code	Title	Organism	%ID	Query rgn.	Target rgn.	Aln. score	Method	Confidence
1	1A5A:A	CRYO-CRYSTALLOGRAPHY OF A TRUE SU...	<i>Syphilium</i>	10.9	42 - 172	120 - 250	54	Local	Low(76%)
2	1JLM	I-DOMAIN FROM INTEGRIN CR3, MN2+ BO...	<i>Hsapilens</i>	11.1	373 - 504	6 - 145	55	Local	Low(70%)
3	1BHO:1	MAC-1 I DOMAIN MAGNESIUM COMPLEX	<i>Hsapilens</i>	11.7	373 - 504	6 - 145	55	Local	Low(67%)
4	1BHO:2	MAC-1 I DOMAIN MAGNESIUM COMPLEX	<i>Hsapilens</i>	11.7	373 - 504	6 - 145	55	Local	Low(67%)
5	1IDN:2	MAC-1 I DOMAIN METAL FREE	<i>Hsapilens</i>	11.7	373 - 504	6 - 145	55	Local	Low(67%)
6	1BHQ:2	MAC-1 I DOMAIN CADMIUM COMPLEX	<i>Hsapilens</i>	11.7	373 - 504	6 - 145	55	Local	Low(67%)
7	1BHQ:1	MAC-1 I DOMAIN CADMIUM COMPLEX	<i>Hsapilens</i>	11.7	373 - 504	6 - 145	55	Local	Low(67%)
8	1IDN:1	MAC-1 I DOMAIN METAL FREE	<i>Hsapilens</i>	11.7	373 - 504	6 - 145	55	Local	Low(67%)
9	1BEO:A	TRP SYNTHASE (D80N-IPP-SER) WITH K+	<i>Syphilium</i>	10.3	24 - 172	102 - 248	53	Local	Low(58%)
10	1IDO	I-DOMAIN FROM INTEGRIN CR3, MG2+ BO...	<i>Hsapilens</i>	10.9	373 - 488	6 - 131	58	Local	Low(55%)
11	2TYS:A	CRYSTAL STRUCTURES OF MUTANT (BETA...	<i>Syphilium</i>	11.5	25 - 174	103 - 252	56	Local	Low(53%)
12	TAOX:A1	DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>Hsapilens</i>	15.0	411 - 504	45 - 148	53	Local	Low(53%)
13	1UBS:A	TRYPTOPHAN SYNTHASE (EC.4.2.1.20) WT...	<i>Syphilium</i>	11.5	42 - 174	120 - 254	65	Local	Low(48%)
14	1BKS:A	TRYPTOPHAN SYNTHASE (EC.4.2.1.20) FR...	<i>Syphilium</i>	10.9	42 - 172	120 - 251	61	Local	Low(36%)
15	1OAK:A	CRYSTAL STRUCTURE OF THE VON WILLE...	<i>Hsapilens</i>	6.7	366 - 504	4 - 150	49	Local	Low(34%)
16	TAOX:B	DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>Hsapilens</i>	10.3	411 - 504	47 - 148	48	Local	Low(31%)
17	1GGC:1	STRUCTURE OF THE COMPLEX OF AN FAB...	<i>Falisse</i>	21.1	277 - 355	10 - 100	62	Local	Low(23%)
18	1FMD:1	FOOT AND MOUTH DISEASE VIRUS TYPE...	<i>Falisse</i>	21.1	277 - 355	10 - 100	62	Local	Low(22%)
19	1ALUQ	A1 DOMAIN OF VON WILLEBRAND FACTOR	<i>Hsapilens</i>	7.0	373 - 504	20 - 159	43	Local	Low(21%)
20	1BWOB	CRYSTAL STRUCTURE OF TYROSINE AMIN...	<i>Tarzel</i>	5.5	463 - 536	178 - 248	57	Local	Low(15%)
21	1TTP:A	TRYPTOPHAN SYNTHASE (EC.4.2.1.20) IN T...	<i>Syphilium</i>	10.4	44 - 172	122 - 251	51	Local	Low(15%)
22	1TTG:A	TRYPTOPHAN SYNTHASE (EC.4.2.1.20) IN T...	<i>Syphilium</i>	10.4	44 - 172	122 - 251	51	Local	Low(15%)
23	1ART	ASPARTATE AMINOTRANSFERASE (EC.2.6...	<i>Ecoli</i>	10.4	405 - 537	116 - 248	52	Local	Low(12%)
24	1GITA	ASPARTATE AMINOTRANSFERASE FROM...	<i>N/A</i>	10.5	405 - 537	116 - 248	45	Local	Low(12%)
25	1EEGA	CRYSTAL STRUCTURE OF YEAST KARYOP...	<i>Sacerevisiae</i>	15.5	35 - 130	238 - 336	49	Local	Low(10%)
26	1EEGB	CRYSTAL STRUCTURE OF YEAST KARYOP...	<i>Sacerevisiae</i>	15.5	35 - 130	238 - 336	49	Local	Low(10%)
27	1ARHA	ASPARTATE AMINOTRANSFERASE, Y225R/...	<i>Ecoli</i>	10.5	405 - 537	116 - 248	46	Local	Low(10%)
28	1ARHB	ASPARTATE AMINOTRANSFERASE, Y225R/...	<i>Ecoli</i>	10.5	405 - 537	116 - 248	46	Local	Low(10%)
29	1EFTA	CRYSTAL STRUCTURE OF THE MOESIN FE...	<i>N/A</i>	18.8	82 - 171	121 - 188	55	Local	Low(10%)
30	1TEF:B	CRYSTAL STRUCTURE OF THE MOESIN FE...	<i>N/A</i>	18.8	82 - 171	121 - 188	55	Local	Low(10%)

GenBank SWISS-PROT PDB

File Edit Search Type Help

FIG. 22B

Aligned Sequence Display

Query details:

P10155: 60 KDA RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO) (RORNP) (SJORGEN SYNDROME TYPE A ANTIGEN (SS-A)).



Tools

Sequences

Ligands

Total selected: 1. View alignment

Reverse Maximised PSI-BLAST results:

Select all
Deselect all

Total hits: 13
Selected: 0

Page 1 of 1

Cluster Detail	Code	Title	Organism	%ID	Query rgn.	Target rgn.	Filter	E value
♦	AAA35532.1	Not given	Hsaplens	99.0	1 - 516	1 - 516	1	0.00E+00
♦	AAF19049.1	Not given	Mmusculus	90.0	1 - 538	1 - 537	1	0.00E+00
♦	AAC53142.1	Not given	Mmusculus	90.0	1 - 538	1 - 538	1	0.00E+00
♦	AAC38001.1	ribonucleoprotein	Xlaevis	78.0	1 - 538	1 - 538	1	0.00E+00
♦	AAB81552.1	Not given	Hsaplens	100.0	1 - 189	1 - 189	1	1.00E-111
♦	CAA98241.1	Not given	Celegans	36.0	7 - 538	54 - 643	2	0.00E+00
♦	AAF10933.1	Not given	Dradiodurans	34.0	7 - 538	15 - 531	2	0.00E+00
♦	AAF45876.1	Not given	Dmelanogaster	17.0	1 - 537	64 - 639	3	1.00E-126
♦	CAB46406.1	hypothetical protein	Scorcolor	26.0	266 - 381	273 - 372	3	6.00E-25
♦	AAC53043.1	telomerase protein-1	Mmusculus	21.0	264 - 378	513 - 627	3	6.00E-05
♦	AAB51690.1	telomerase protein component 1	Rnorvegicus	23.0	284 - 375	517 - 627	3	2.00E-04
♦	AAC46501.1	telomerase component p80	Tthermophila	27.0	264 - 329	361 - 425	3	7.00E-04
♦	AAC51107.1	telomerase-associated protein TP-1	Hsaplens	25.0	264 - 327	507 - 669	3	0.00E-03

GenBank SWISS-PROT PDB

FIG. 23

AIEye output (January 2, 2001 1:29 PM)

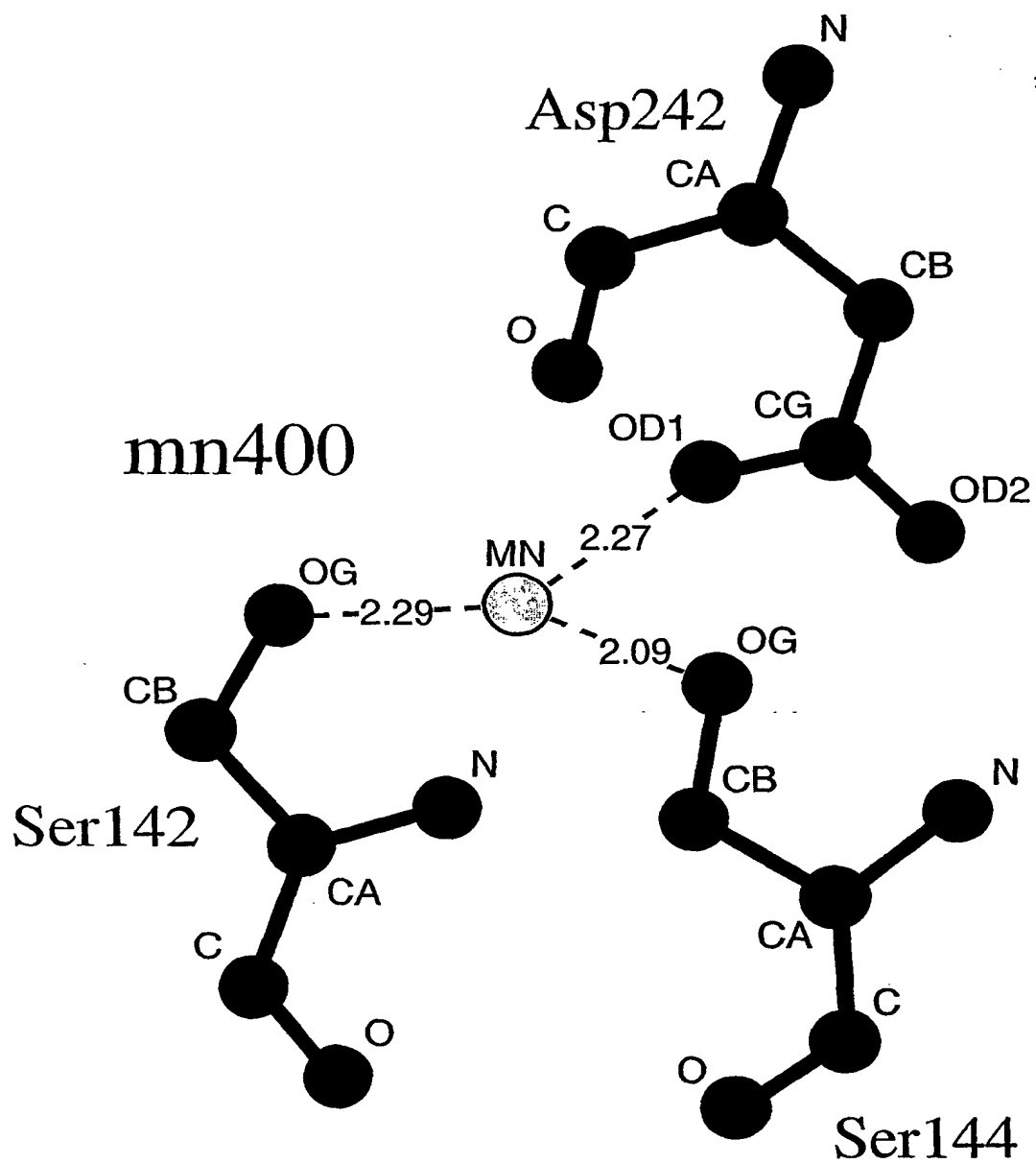
36/40

P101SS 1JM	10	20	30	40	50	60	70	80
	MEESVNMQPLNEKQ	IANSQDGYVWQVTDNNRLHRLFCGSEGGTY	YI	KEQKLGLENAEAL	RLIEDGRGCEVI	QEKSF		
P101SS 1JM	90	100	110	120	130	140	150	160
	SQEGRTTKQEPMLFALA	ICSQCSDISTKQAAFKAVSEVCR	IP	THLFTFIQFKKDLKESMKCGMWGRAL	RKA	IADWYNEKG		
P101SS 1JM	170	180	190	200	210	220	230	240
	GNALALAVTKYKQRNGW	SHKDLLRLSHLKPSSGLA	IV	TKYITKGWKEVHELYKEKALS	YET	EKLLKYLEAVEKVKRTKD		
P101SS 1JM	250	260	270	280	290	300	310	320
	ELEVHLIEEHLVREHL	LTNHLKSKEVWKALLQEMPLT	ALLRN	LGKMTANSVLEPGNSEVSLVCEKLCNEKLLKKARIH				
P101SS 1JM	330	340	350	360	370	380	390	400
	PFHILIALETYKTGH	GLRGKLGKWRPDEEIL	KALDA	AFYKTFKTVEPTGKRFL	--L--	AVDVASAMNQRLVLSILNASTVA		
P101SS 1JM	410	420	430	440	450	460	470	480
	A-----AMCMVVTR	TEKDSYVVA	FSD	EMVPCPVTTDMTLQQVLNAMSQI	-	PAGGTDCSLPMIWAQKTNT	---	PADVFIYF
P101SS 1JM	490	500	510	520	530	540	550	560
	NEQLKSKTLFSLNQ	SEEFRIHFT	FK	EQNNPNPRSLVKPITQLLGR	TH	TATGIRKVVRELFNITNGARKNAFKILVVI		
P101SS 1JM	570	580	590	600	610	620	630	640
	TONETFAGGVHPA	I	ALREYRKKMDI	PAKLIVCGMTSNGF	-----	TIADPDDRG-MLDMCGFDTGALDVI	RNFTLDM	I
P101SS 1JM	650	660	670	680	690	700	710	720
	TOGEKFGDPLGY	EDVIPE	-ADREGVIRYVIGVGD	AFRSEKSRQELNT	IASKPP	PRDHVFQYNNFE	--	ALKTIQNQLREKIF

P101SS
1JM A

37/40

FIG. 24A



38/40

FIG. 24B



39/40

FIG. 25

AlEye output (January 4, 2001 3:18 PM)

P10155	MEESVN	10	20	30	40	50	60	70
AAF19049.1	MEGSAN	10	20	30	40	50	60	70
AAC38001.1	MEATMD	10	20	30	40	50	60	70
CAA98241.1	MADELNEFOEAGNFNEEALMRLSNVCA	10	20	30	40	50	60	70
	RLRMOMLES	10	20	30	40	50	60	70
	SDVEITVVDGELKRVPRQMEKVKDQGV	10	20	30	40	50	60	70
	ENNAGGFVFP	10	20	30	40	50	60	70
P10155	VTDMNRLHRRFLCFGSEGGTTYI	80	90	100	110	120	130	140
AAF19049.1	VTDMNRLHRRFLCFGSEGGTTYI	80	90	100	110	120	130	140
AAC38001.1	VTDMNRLHRRFLCFGSEGGTTYI	80	90	100	110	120	130	140
CAA98241.1	VTDMNRLHRRFLCFGSEGGTTYI	80	90	100	110	120	130	140
	KEQKLGLENAEALIRLIEDGRGCEVI	80	90	100	110	120	130	140
	QEI	80	90	100	110	120	130	140
	KSFSGEGRTTKQEPMLFALACSQCS	80	90	100	110	120	130	140
	SD	80	90	100	110	120	130	140
P10155	ISTKQ	160	170	180	190	200	210	220
AAF19049.1	INTKQ	160	170	180	190	200	210	220
AAC38001.1	IKTKQ	160	170	180	190	200	210	220
CAA98241.1	HDTTKTECPMLNAYS	160	170	180	190	200	210	220
	DIPEVORTPTHLEFVDYQCTISEST	160	170	180	190	200	210	220
	KAGGAKSSTGWGRSMNAISKWYT	160	170	180	190	200	210	220
P10155	EKGGMALALAVTKYKQ	240	250	260	270	280	290	300
AAF19049.1	EKGGMALALAVTKYKQ	240	250	260	270	280	290	300
AAC38001.1	EKGGMALALAVTKYKQ	240	250	260	270	280	290	300
CAA98241.1	EKGGMALALAVTKYKQ	240	250	260	270	280	290	300
	NRNGWSHKDLLRLSH	240	250	260	270	280	290	300
	---	240	250	260	270	280	290	300
	LKPSSEGLA	240	250	260	270	280	290	300
	---	240	250	260	270	280	290	300
	VTKYITKG	240	250	260	270	280	290	300
	---	240	250	260	270	280	290	300
	WKEVHEL	240	250	260	270	280	290	300
P10155	YKEKALSV	320	330	340	350	360	370	380
AAF19049.1	YKEKALSV	320	330	340	350	360	370	380
AAC38001.1	YKEKALSV	320	330	340	350	360	370	380
CAA98241.1	YKEKALSV	320	330	340	350	360	370	380
	ETEKLLKYLEAVEKVKRTKDELEV	320	330	340	350	360	370	380
	---	320	330	340	350	360	370	380
	EHRLVREHLLTNHLKSKEVWKALLO	320	330	340	350	360	370	380
	---	320	330	340	350	360	370	380
	EMPLT	320	330	340	350	360	370	380
P10155	WDKKALKLPYTEQL	400	410	420	430	440	450	460
AAF19049.1	WDKKALKLPYTEQL	400	410	420	430	440	450	460
AAC38001.1	WDKKALKLPYTEQL	400	410	420	430	440	450	460
CAA98241.1	WDKKALKLPYTEQL	400	410	420	430	440	450	460
	KEEQSRALNLVEAYLKLKNEQSEEV	400	410	420	430	440	450	460
	---	400	410	420	430	440	450	460
	VAAIKKHGLVREHLPTTSLNSKL	400	410	420	430	440	450	460
	---	400	410	420	430	440	450	460
	WETLFDVSMPT	400	410	420	430	440	450	460
P10155	ALLRNLGKMTANSVLEPGNSEV	480	490	500	510	520	530	540
AAF19049.1	ALLRNLGKMTANSVLEPGNSEV	480	490	500	510	520	530	540
AAC38001.1	ALLRNLGKMTANSVLEPGNSEV	480	490	500	510	520	530	540
CAA98241.1	ALLRNLGKMTANSVLEPGNSEV	480	490	500	510	520	530	540
	SLCEKLSNEKLLKKARIHPFFHIL	480	490	500	510	520	530	540
	---	480	490	500	510	520	530	540
	ALETYRAGHGLRGKLRWPDKEIL	480	490	500	510	520	530	540
	---	480	490	500	510	520	530	540
	KALDAAF	480	490	500	510	520	530	540
P10155	ALLRNLGKMTANSVLEPGNSEV	560	570	580	590	600	610	620
AAF19049.1	ALLRNLGKMTANSVLEPGNSEV	560	570	580	590	600	610	620
AAC38001.1	ALLRNLGKMTANSVLEPGNSEV	560	570	580	590	600	610	620
CAA98241.1	ALLRNLGKMTANSVLEPGNSEV	560	570	580	590	600	610	620
	SLCEKLSNEKLLKKARIHPFFHIL	560	570	580	590	600	610	620
	---	560	570	580	590	600	610	620
	ALETYRAGHGLRGKLRWPDKEIL	560	570	580	590	600	610	620
	---	560	570	580	590	600	610	620
	KALDAAF	560	570	580	590	600	610	620
P10155	ALLRNLGKMTANSVLEPGNSEV	640	650	660	670	680	690	700
AAF19049.1	ALLRNLGKMTANSVLEPGNSEV	640	650	660	670	680	690	700
AAC38001.1	ALLRNLGKMTANSVLEPGNSEV	640	650	660	670	680	690	700
CAA98241.1	ALLRNLGKMTANSVLEPGNSEV	640	650	660	670	680	690	700
	SLCEKLSNEKLLKKARIHPFFHIL	640	650	660	670	680	690	700
	---	640	650	660	670	680	690	700
	ALETYRAGHGLRGKLRWPDKEIL	640	650	660	670	680	690	700
	---	640	650	660	670	680	690	700
	KALDAAF	640	650	660	670	680	690	700

40/40

FIG. 25 (contd.)

AIEye output (January 4, 2001 3:18 PM)

P10155	470	480	490	500	510	520	530	540	
AAF19049.1	YKTFKTVEPTGKRFLLA	VDSASMNQORVLGS	ILNASTVAAAMCMVV	TRTEKDSYVVAFSDEM	VPCPVTTDMTLQQVLM				
AAC38001.1	YKTFKTVEPTGKRFLLA	VDSASMNQORVLGS	ILNASTVAAAMCMVV	TRTEKDSYVVAFSDEM	VPCPVTTDMTLQQVLT				
CAA88241.1	YKAFVNA	PPTGKRYCLALDVSG	SMTSRVSSPLSCRE	ATGMSLINLHNEAEV	RCVAFCDKLT	TEL	PFTKDWKIGQVND		
P10155	550	560	570	580	590	600	610	620	
AAF19049.1	AMSQIPAGGTDCSLP	MIWAQKTNTPADVF	FIVFTDNETFAGGV	HPAIALREYRKKMDI	-PAKLI	VCGMTSNGFTIADPD			
AAC38001.1	AMNKVPAGNTDCSLP	MIWAQKTGTADVF	FIVFTDNETFAGGV	HPAIALREYRKKMDI	-PAKLI	VCGMTSNGFTIADPD			
CAA88241.1	KMSDITMGSTDCAL	PMLWAQKTNTAAD	I	FIVFTDCETNVED	VHPATALKQYREKMG	I-PAKLI	VCAMTSNGFSIADPD		
	YVNNLDFGR	TDGCLPMTWATENN	LKFDVFII	YTDNDTWAGEI	HPFEAIKKYREASG	I	HDAKVI	VMAMQAYDYSIADPS	
P10155	630	640	650						
AAF19049.1	DRGMLDMCGFD	TGALDVIRNFTLDM	I						
AAC38001.1	DRGMLDMCGFD	TGALDVIRNFTLDM	I						
CAA88241.1	DRGMLD	ICGFD	SGALDVIRNFTL	DLI					
	DAGMLD	ITGFD	SAVPQIVHEF	VTGKI					